Nucleic Acids Research, 1993, Vol. 21, No. 16 3683–3689

I-*Sce* III an intron-encoded DNA endonuclease from yeast mitochondria. Asymmetrical DNA binding properties and cleavage reaction

M.Schapira, C.Desdouets+, C.Jacq* and J.Perea
Laboratoire de Génétique Moléculaire, CNRS URA1302, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris Cedex 05, France

Received May 11, 1993; Revised and Accepted July 6, 1993

ABSTRACT

We have previously discovered the new intron-encoded endonuclease I-*Sce* III by expressing, in *E.coli*, the ORF contained in the third intron of the yeast mitochondrial COX I gene. In this work, we analyzed the *in vitro* properties of partially purified I-*Sce* III and found that it is a very specific DNA endonuclease, tolerating relatively few base changes in its 20 base pair long target site. I-*Sce* III should be a useful molecular tool to analyze the structure of large genomes. Interestingly, I-*Sce* III is the first P1-P2 DNA endonuclease for which DNA binding properties could be analyzed by band-shift experiments. Clearly, the cleavage products corresponding to the upstream A3 exon and to the downstream A4 exon could compete with the substrate A3-A4 in forming a DNA-protein complex. However, the A3 exon competes more efficiently than the downstream A4 product. The cleavage of the two DNA strands is also asymmetric the top strand (non-transcribed strand) is cleaved faster than the bottom strand, a property found under various experimental conditions. These findings suggest that this intron-encoded DNA endonuclease may have a role in the RNA splicing process of the intron.

INTRODUCTION

The majority of group I intron-encoded proteins whose function is known act either as maturase to promote splicing or as specific DNA endonuclease to initiate intron mobility (1–3). In yeast mitochondrial introns, where these two types of activities were first discovered, the intron-encoded proteins belong to a major class of homologous proteins sharing two highly similar dodecapeptides called LAGLI-DADG or P1-P2 (4, 5). Interestingly, other members of this family, which have DNA endonuclease activities, are not intron encoded in mitochondrial (6, 7) or nuclear genomes (8). Recently, such gene encoding P1-P2 endonucleases have been found inserted directly in protein-coding sequences from where they are likely to be excised by an intriguing protein-splicing mechanism (9–12),(13) for a review.

Concerning group I introns, there is mounting evidence to support the idea that intron ORFs coding for DNA endonucleases were acquired by the intron in which they found a genetic refuge through the ability of the intron to splice (14–16). Molecular adaptation of these ORFs to their new environment would have followed for the benefit of both the intron and the intron-encoded protein. The intron ORF, in most of the group I introns, is in phase with the upstream exon and the endonuclease has acquired a target specificity for the sequences that flank the site of intron insertion. In some cases the intron-encoded protein became completely devoted to the intron splicing process by losing its DNA endonuclease activity and developing an RNA maturase activity. Close relationships between these two activities have been experimentally assessed (17) in the case of two homologous intron-encoded proteins from the yeast mitochondrial genome one controlling the splicing of the two introns while the other has a DNA endonuclease activity which can stimulate the propagation of its own intron. The yeast mitochondrial genome is in fact one of the most completely understood genetic system in terms of function of group I intron-encoded proteins. Some points, however, remain obscure. First, some intron-encoded proteins have still unknown functions and it would be interesting to know whether they belong to one of the two types of activities mentioned above. Second, the molecular properties of either DNA endonuclease or RNA maturase are still poorly understood and it is still difficult to know their exact function in either the intron mobility or the RNA splicing processes. Only in the first case it seems clear that the double-strand break introduced by the DNA endonuclease in the intronless allele is the first step of the gene conversion event. The subsequent role of the P1-P2 DNA endonuclease in the double-strand-break repair process leading to the synthesis of a new copy of the intron sequence in its new site (homing site, (18)) remains to be established.

In this work, we have discovered in the mitochondrial genome of *Saccharomyces cerevisiae* (strain 777-3A), a DNA endonuclease activity associated with the protein coded in the

* To whom correspondence should be addressed

+ Present address: Hôpital Necker, 149 rue de Sèvres 75015 Paris, France

third intron of the cytochrome oxidase subunit I gene (COX I). We expressed this protein in *E.coli* after adapting its coding sequence to the bacterial genetic code and characterized its *in vitro* properties. This endonuclease is especially interesting since it is the first protein of this class with which it has been possible to directly study DNA binding properties by band-shift analyses. We also analyzed the specificity of the protein for its target site and the sequential cleavage of the two strands. These experiments have been conducted to better understand the role of the intron-encoded proteins in their adaptation to the genomic environment.

MATERIALS AND METHODS

Universal code adaptation of the I-*Sce* III coding sequence

A 1800 bp BglII-BamHI fragment containing the ORF of third intron of the COX I gene (oxi3) from the yeast strain 777-3A was cloned in the Bam HI site of pBlueScriptKS$^+$ (Stratagene). A recombinant plasmid pAI3NC containing the insert in an orientation in which the non-transcribed strand is ligated to the plus strand of the vector (pAI3NC) was used as a substrate for the directed mutagenesis. A set of mutagenic oligonucleotides was designed from the published sequence of the COX I gene of the strain D273-10B (19) in order to modify 22 codons and to adapt the I-*Sce* III coding sequence to the standard genetic code. Single stranded DNA was obtained from TG1 transformed with pAI3NC by infection with the helper phage MK07. 15 pmol of each mutagenic oligonucleotide was phosphorylated with polynucleotide kinase and mixed with about 1μg of ssDNA (in 20 mM Tris/HCl pH7, 5, 10 mM $MgCl_2$, 50 mM NaCl and 1 mM dithiothreitol in a final volume of 20μl). Hybridization was done by heating at 92°C for 5 min and allowing to cool slowly to room temperature. 20 μl of Elongation mix (20 mM Tris/HCl pH7, 5, 10 mM $MgCl_2$, 10 mM ATP, 10 mM Dithiothreitol, 2 mM each dNTP, 1 unit of ligase and 3 units of klenow enzyme) was added and the resulting reaction was incubated 2h at 37°C.

Oligonucleotides were removed with GeneClean (BIO101) and the DNA was collected in 50 μl of TE. 10 μl of this DNA was used in a PCR with 20 pmol of the two oligonucleotides NC (CTCGAACGTCAGCCTGAATC) and CC (TTGGCTGAA-GGCTACGGTAG). This pair of oligonucleotides amplify selectively the strand which has been synthesized by inserting the mutagenic oligonucleotides.

The PCR product was electrophoresed, purified from the agarose gel and cloned in pBlueScript KS+ plasmid. Several clones were sequenced using the mutagenic oligonucleotides as primers.

Sequences corresponding to three of the oligonucleotides were never modified by this approach. In order to investigate this failure we used the mutagenic oligonucleotides to sequence a wild type I-*Sce* III gene from the strain 777-3A. We discovered 16 polymorphic changes between the strain D273-10B (19) which was used to design the mutagenic oligonucleotides and 777-3A (source of DNA). Some of these base changes were in the sequence corresponding to the mutagenic oligonucleotides which probably did not hybridize to their target. New oligonucleotides were resynthesized with the corrected sequence and incorporated by the Kunkel method of targeted mutagenesis (20).

Three other polymorphic changes corresponding to the strain D273−10B were included in the mutagenic PCR experiment a Ser_{28} to Thr_{28}, a Thr_{82} to Pro_{82} and a Asn_{286} to Lys_{286}. The corresponding protein had no DNA endonuclease activity.

The DNA endonuclease activity of I-*Sce* III was found when Pro_{82} was changed into Thr_{82} as it should be in the strain 777-3A. The sequence of I-*Sce* III is presented in figure 1.

Expression and purification of I-*Sce* III

The adapted gene coding for the protein was cloned in the Bam HI site of the bacterial expression vector pDR540 which contains the conditional tac promoter (Pharmacia). A two liter culture of transformed *E.coli* (strain XL1) was grown for 2 hours at 37°C. When the OD_{600} reached 0.6, I-*Sce* III synthesis was induced by adding IPTG to a final concentration of 3mM. Cells were grown for an additional 2 hours, harvested by centrifugation at 8500 rpm for 20 minutes at 4°C and stored at −80°C until use. The cell pellet was then suspended in 10 ml PBS buffer (20 mM H_3PO_4-2 mM EDTA-150mM NaCl-2mM β-mercaptoethanol-8 μM PMSF-pH 7.3), freezed in liquid nitrogen and heated to 37°C twice. After the addition of 1ml 10% NP40, the cells were sonicated, PMSF was added to a final concentration of 16 μM and the bacterial lysate was centrifuged at 4°C for 20 minutes at 15000 rpm.

All purification steps were performed at 4°C The supernatant was further centrifuged at 4°C for 20 minutes at 15000 rpm and directly loaded on a 8 ml phosphocellulose column (Pharmacia HR 10/10) previously equilibrated with PBS buffer. The resin was washed with 10 column volumes of PBS and the protein was eluted, with a linear gradient, at a ionic strength of 0.8 M NaCl. The active fractions were pooled and desalted to a 10 mM NaCl final concentration on a Pharmacia HR 10/10 desalting column. One third of the preparation was immediately loaded on a 1 ml cation exchanger mono S column (Pharmacia HR 5/5) previously equilibrated with PBS 10 mM NaCl. The active fractions were recovered at a 0.6 M NaCl concentration and were stored at 4°C. 3000 UA of I-*Sce* III were usually obtained from a two liter culture; one UA (unit of activity) being defined as the amount of enzyme necessary to digest 1 μg of a target containing plasmid in 30 minutes at 37°C.

I-*Sce* III target mutagenesis

Two degenerate oligonucleotides 1.—GATCGGTTTTGGTAA-CTATTTATTAC and 2.—CCAAAACCATTGATAAATAAT-GCTAG were synthesized with a mixture of phosphoramidites (92.5% of the base corresponding to the wild type sequence and 2.5% of each other three bases), hybridized and cloned in the BamHI site of pBluescript KS$^+$. We sequenced 75 clones and we found 35 wild type, 29 single mutants, 10 double mutants and a triple mutant. From the 29 single mutants, 26 were different and were checked as substrates for the I-*Sce* III endonuclease activity as well as the 10 double mutants.

Cleavage assay on yeast genome

Introduction of the I-*Sce* III target site in chromosome III A 2870 bp EcoRI BamHI fragment from chromosome III (site EcoRI 145929; site BamHI 148799 (21)) was cloned in pBluescript KS$^+$. The resulting plasmid was used to clone a 22 bp long target sequence in the unique BamHI site (pDEL5). This plasmid was integrated in the chromosome III of W303-1B by linearisation at the SnaBI unique site, transformation and selection of the URA3 marker. Digestion of the yeast genome and pulsed field electrophoresis analysis 50 μl agarose plugs containing 5.10^7 yeast cells (disrupted or wild type diploïd strains) were prepared as described by Schwartz and Cantor (22), washed and

equilibrated three times for one hour in 1 ml A_0 buffer (33mM Tris Acetate (pH 7.9), 66 mM K Acetate, 0.5 mM DTT, 100 μg/ml BSA) at 4°C. The plugs were then incubated for three hours at 4°C in 250 μl A_0 buffer plus 30 μl of semi-purified I-*Sce* III preparation to let the enzyme diffuse into the agarose. 30 μl of 100 mM $MgCl_2$ was then added and the mixtures were heated to 37°C for 40 minutes. The plugs were then equilibrated in the running buffer (100 mM Tris (pH 8), 100 mM Boric acid, 0.2 mM EDTA) and inserted into the wells of a 1% Seakem agarose gel. The migration was performed at 120 V for 36 hours with a pulse time of 100 seconds at 10°C ('Pulsaphor', Pharmacia).

Blot and hybridization with a labelled probe The gel was treated for 35 minutes with 0.25 N HCl, for 35 minutes with 0.4 N NaOH and the DNA was blotted for 3 hours under vacuum on an Hybond-N^+ membrane. 0.2 μg of the recombinant pDEL5 plasmid with the A3-A4 inserted sequence was labelled by random priming (Boehringer kit) and was then purified on a 2 ml Sephadex G50 column equilibrated with TE. The resulting probe was heated for 5 minutes to 90°C and used for an overnight hybridyzation in Church buffer (23) at 65°C. The membrane was then washed three times in 2×-0.5×-0.1×SSC buffer 0.1% SDS and autoradiographed on a Kodak X-omat film.

Band-shift assays

The oligonucleotides used as probes or as competitors for bandshift assays were purified on a 20% polyacrylamide (191) denaturing gel and, for the former, labelled with T4 polynucleotide kinase. After the labelling reaction, the mixture was heated to 68°C for 10 minutes and hybridization was carried out by adding an equal amount of the complementary strand. The kinase was removed from the reaction mixture by phenol extraction and the resulting probe was precipitated and dissolved in water to a final concentration of 10^{-8} M. Binding reactions were carried out on ice in 20 μl binding buffer (33 mM Tris Acetate (pH 7.9), 10 mM Mg Acetate, 66mM K Acetate, 0.5 mM DTT, 250 μg/ml BSA). Salmon sperm DNA was used as non specific competitor to a final concentration of 25 μg/ml. 2μl (i.e 20. 10^{-9} mol.) of probe and adequate amounts of specific competitor were added and preincubated for 10 minutes on ice before the addition of 2 μl (0.5 u) of semi-purified enzyme. After another 10 minutes on ice, the binding reaction was complete. Free DNA and protein-DNA complexe were resolved on a native 6% polyacrylamide gel (291) run in 0.5×TBE at 4°C. To generate double-stranded oligonucleotides containing derivatives of I-*Sce* III target site, the following oligonucleotides were synthetised:

A3A4 (Top): 5'-GATCGGAGGTTTTGGTAACTATTTATTACCA-3'
A3A4 (Bottom): 5'-GATCTGGTAATAAATAGTTACCAAAACCTCC-3'
A3 (Top): 5'-AATTGCTTTAATTGGAGGTTTTGGTAAC-3'
A3 (Bottom): 5'-CCAAAACCTCCAATTAAAGC-3'
A3al3 (Top): 5'-AATTGCTTTAATTGGAGGTTTTGGTAACCAAAAAA-GATATG-3'
A3al3 (Bottom): 5'-AATTCATATCTTTTTTGGTTACCAAAACCTCCAAT-TAAAGC-3'
al3A4 (Top): 5'-TAATAAAATGAACTATTTATTACCATTAATAA-TTGGAGC-3'
al3A4 (Bottom): 5'-GCTCCAATTATTAATGGTAATAAATAGTTCAT-TTTATT
A-3' A4 (Top): 5'-TATTTATTACCATTAATAATTGGAGC-3'
A4 (Bottom): 5'-AATTGCTCCAATTATTAATGGTAATAAATAGTTA-3'
Flag (Top): 5'-GATCATGGATTACAAGGATGACGATGATAAAG-3'
Flag (Bottom): 5'-GATCCTTTATCATCGTCATCCTTGTAATCCAT-3'

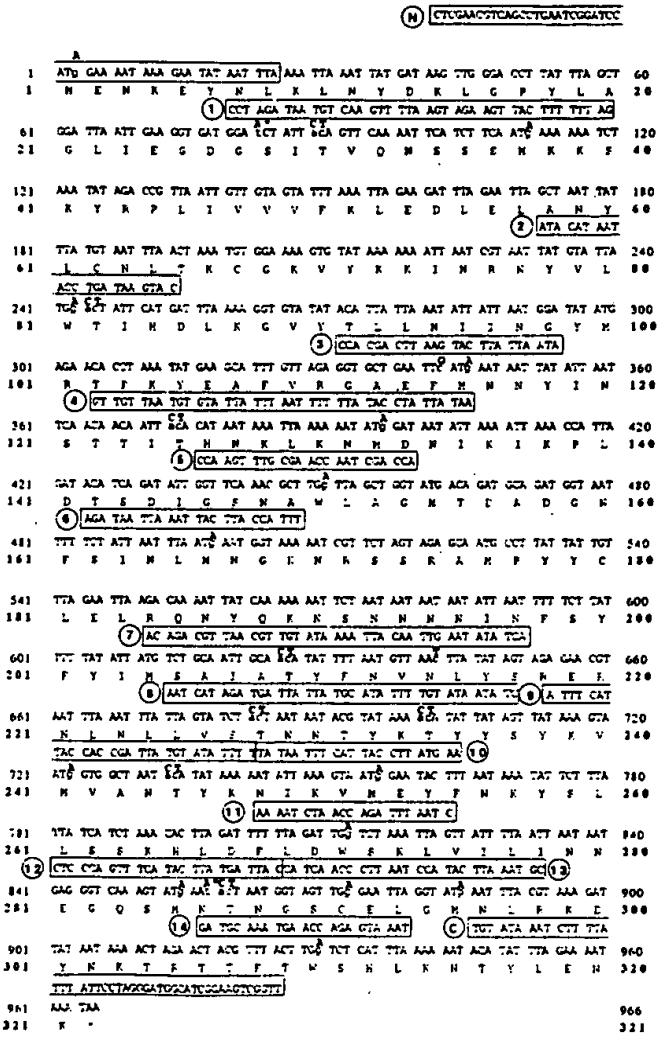


Figure 1. Sequence of the universal code equivalent of the al3 intronic ORF. The sequence of the non transcribed strand is given together with the sequence of the protein as it is translated in *E.coli*. The bases indicated above the nucleic acid sequence correspond (bold letters) to the mitochondrial sequence (strain 777-3A) before it was mutagenized. Boxed sequences show the sequences of the oligonucleotides used either to adapt the mitochondrial coding sequence to the *E.coli* genetic code (1 to 14) or to be used as PCR primers to select the mutated strand with the initiator and terminator codons.

RESULTS

Construction of a universal code equivalent gene from the al3 intronic ORF

The mitochondrial genome of *Saccharomyces cerevisiae* contains a COX I gene interrupted by either group I or group II introns. In the strain D273-10B, the third intron in this gene is a group I intron which contains an open reading frame of 347 codons in phase with the upstream exon (19). The putative translation product of this ORF contains the two conserved motifs P1 and P2 (also called LAGLI and DADG (4, 5)). Taking advantage of our previous knowledge of the properties of similar intronic proteins (24) we presumed that the expression of a 321 aminoacid long polypeptide, corresponding to the C terminal part of the intronic protein, should be sufficient to observe the putative

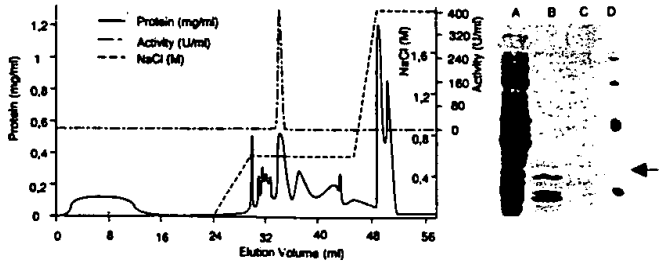


Figure 2. Partial purification of I-*Sce* III on a Mono-S fast colum. The *E.coli* extracts containing artificially produced I-*Sce* III were chromatographed on phosphocellulose P11 as described in the 'Materials and Methods' and the active fractions were pooled and chromatographed on a Mono-S fast column (Pharmacia). 40% of the enzymatic activity was recovered after the Mono-S chromatography (left). SDS-PAGE analysis of fractions containing I-*Sce* III (right). Protein samples were denatured in SDS-PAGE sample buffer and electrophoresed on a 15% SDS-polyacrylamide gel. The gel was fixed and comassie-blue stained. Lane a, *E.coli* extract, 8 μg; lane b, phosphocellulose fractions, 10 μg. (60 I-*Sce* III units). Lane c, Mono-S, 5 μg. (110 I-*Sce* III units). The protein standards used and their molecular weight (kDa) are carbonic anhydrase, 28,5; ovalbumine, 43,7; bovine serum albumin, 70,6. The arrow indicates the position of I-*Sce* III.

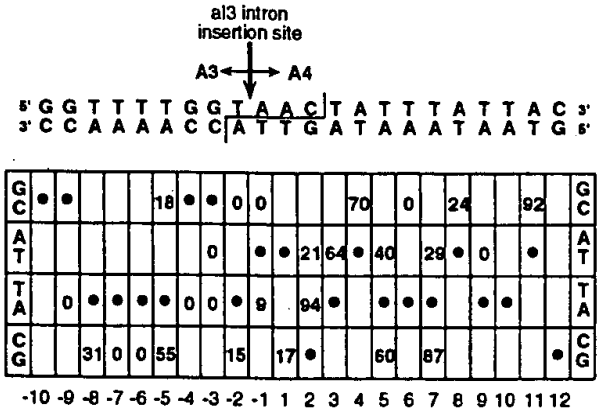


Figure 3. In vitro analysis of the effects of point mutations in the target site of I-*Sce* III.The figure presents a synoptic set of site-directed mutations distributed throughout the A3-A4 junction sequence recognized and cut by I-*Sce* III. Cleavage assays were conducted in conditions where about 80% of the wild type sequence was cut. The values indicated in the cases of the 26 single mutants analyzed represent the percentage of cleaved material, taking as reference the cutting ratio found in wild type.

intron-encoded DNA endonucleolytic activity. With this in mind, we modified all the mitochondrial codons which have a different meaning in the *E.coli* genetic code. Thus, 8 CTN codons were changed to ATN, 9 ATA codons were changed to ATG codons and 5 TGA codons were changed to TGG. This was done with the help of the synthetic oligonucleotides depicted in figure 1. A few other base changes were introduced (figure 1) to create useful restriction sites or to maintain polymorphic differences between the strain D273-10B previously sequenced (19) and the strain 777-3A from which we extracted the DNA. The sequence of the engineered coding sequence (figure 1), corresponds mainly to the wild type sequence of the strain 777-3A. Only two amino acid residues from the engineered sequence are specifically found in the published sequence of the strain D273-10B (see Materials and Methods). We noted that replacement of the Thr_{82} (777-3A)

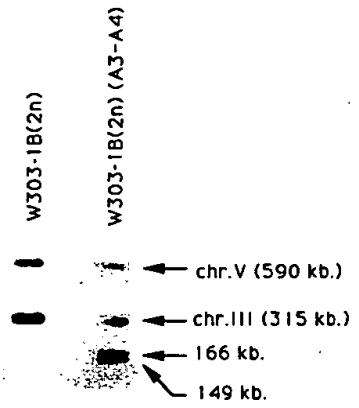


Figure 4. Specific cleavage of the yeast chromosome III by I-*Sce* III. A synthetic form of the I-*Sce* III target site was inserted, by homologous recombination, in the known sequence of chromosome III (see Materials and Methods).The DNA of the transgenic strain thus created was incubated with a purified fraction of I-*Sce* III as described in Materials and Methods. DNA was electrophoresed on 1% agarose gel at 170 V and 10°C on a Pharmacia-LKB 2015 pulsaphor apparatus for 24 hrs using 100 sec pulse time. The DNA was transferred to Hybond N^+ (Amersham) membrane for hybridization with a probe specific of both chromosomes III and V.

by Pro_{82} (D273-10B), abolished the DNA endonuclease activity. Whether this reflects a radical difference between the two strains or simply some sequencing discrepancies remains to be checked. Finally, 5' and 3'-terminal oligonucleotides were made (figure 1) to introduce restriction sites and to be used as PCR primers. The approach followed to introduce the different mutations is described in the section 'Materials and Methods'. Briefly, we took advantage of the PCR method to introduce several base changes in the same round of mutagenesis. Although not perfect (a few base changes had to be introduced by more classical methods) this method considerably sped up the engineering of the gene. The DNA sequence thus obtained could be inserted in a bacterial expression vector to produce and purify the corresponding protein.

Partial purification and in vitro properties of the *E.coli* made intron-encoded DNA endonuclease I-*Sce* III

We previously showed (25) that the engineered ORF coding for the 321 amino acid long protein placed under the control of the tac promoter in the bacterial plasmid pDR 540 (Pharmacia), conferred to the corresponding bacterial extracts a specific DNA endonucleolytic activity. This activity was found to be able to cleave a 20 base pair long DNA sequence corresponding to the fusion of the exons A3 and A4 flanking the al3 intron. We then decided to purify the I-*Sce* III protein to better analyze its properties. The extraction procedure is described in materials and methods. Interestingly, overproduction of I-*Sce* III in *E.coli* did not significantly affect the bacterial growth and large amounts (about 2000 units from one liter culture) of the protein could be obtained. The purification procedure is similar to that of similar DNA cleaving proteins. Phosphocellulose chromatography was followed by a Mono-S column (Figure 2). Active fractions were pooled and chromatographed on a Sephacryl S-200 gel filtration column. The endonuclease activity eluted as a single peak corresponding to a protein with an apparent molecular weight

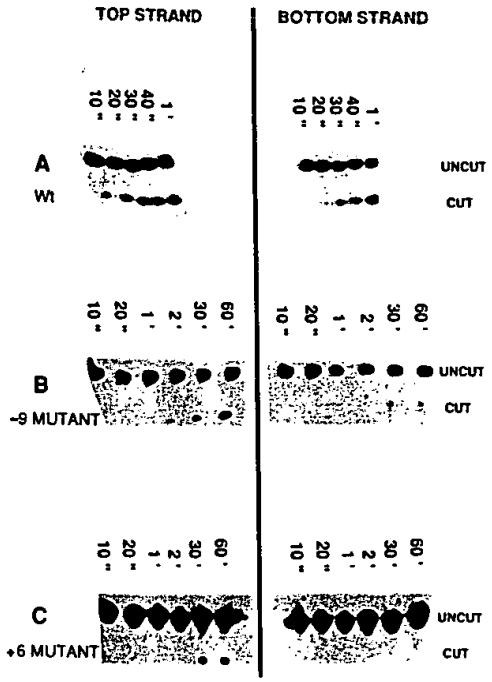


Figure 5. Asymmetrical cleavage at the I-*Sce* III target site. This panel compares the kinetics of I-*Sce* III cleavage at the top and bottom strands for either the wild type sequence (A) or mutated sequences (B) at position −9, GC to TA; (C) at position +6, TA to GC; see Fig.3). The experimental conditions (enzyme preparations, substrates and enzyme concentrations) are the same in the different cases. Labelling of either the top or bottom strand were carried out in the same conditions (Materials and Methods) and aliquots of the reaction mixture with I-*Sce* III were taken at different times (indicated above the different lanes) and analyzed by electrophoresis on denaturing acrylamide gels.

of 34kDa (Stokes radius of 29 A°). This figure is close to the molecular weight of 37kDa calculated from the amino acid composition deduced from the DNA sequence; thus, I-*Sce* III appears as a monomeric globular protein. This feature is reminiscent of the monomeric structure of the intronic protein I-*Sce* I (26) but is at variance with the apparent homodimeric structure of I-*Sce* II (27).

The active endonuclease fractions obtained after this partial purification were free of other non specific endo or exonuclease activities (see figure 2) and could thus be used to study the properties of I-*Sce* III.

The effects of pH, ionic strength, and temperature were analyzed to optimize the assay conditions.The endonuclease activity is limited at low pH values and increases to reach a maximum at pH 8. Our results show that magnesium is absolutely required (optimum concentration = 10 mM) and that the reaction is inhibited by monovalent cations. Thus KCl concentrations above 500 mM dramatically reduce the endonuclease activity.Variations in the reaction temperature between 25 and 50°C have no important effects on the endonuclease activity. Initial rates of the cleavage reaction by I-*Sce* III were determined from assays using various concentrations of substrate. This allowed us to determine a Km value of 6 10^{-9}M at pH 8. All these properties are strikingly similar to those previously found for the protein I-*Sce* I (26). Clearly, the main difference between the two proteins is in their substrate specificity.

Substrate specificity. *In vitro* cleavage of different mutated A3-A4 recognition sites

We have previously observed that a 20 base pair long DNA sequence overlapping the A3-A4 splice junction could be recognized and cleaved by I-*Sce* III. We constructed 26 single mutations (see Materials and Methods) within the A3-A4 recognition site and we introduced them into a pUC vector to analyze their cleavage properties by a purified form of the enzyme I-*Sce* III. The summary of the properties of mutant DNA substrate cleavage is presented in Figure 3. It is interesting to note that at least nine base changes at different positions (−9, −7, −6, −4, −3, −2, −1, +6, +9) in the cleavage site block cleavage substantially (to less than 5% of the wild-type substrate level). This result suggests that I-*Sce* III has a more stringent substrate specificity than I-*Sce* II for which no more than three to five highly critical positions have been identified (28, 29). This observation is in agreement with the fact, mentioned above, that contrary to I-*Sce* II (24), I-*Sce* III can be produced in *E.coli* without any damage to the cell. Also of interest is the fact that the critical positions are not equally distributed along the target site. Six such positions are located in the upstream exon A3, whereas only three of them are located in the downstream exon. As already noted in the case of I-*Sce* II, these critical positions are mostly localized in the first or second codon positions thus giving to the protein more possibilities to promote the horizontal transfer of the intron (discussed in (28)).

The high specificity of I-*Sce* III prompted us to assess its utility as a reagent for physical mapping of complex genomes. Clearly neither small genomes like phage T7 or *E.coli* genomes nor the wild type *Saccharomyces cervisiae* nuclear genome are cleaved by I-*Sce* III. In this last case, we have confirmed that this is due to a lack of appropriate cleavage site by inserting the A3-A4 junction sequence in a precise localization of the nuclear genome and cleaving the corresponding chromosome III at the expected position (Figure 4). Such an approach has already been used to demonstrate the utility of these enzymes to map complex genomes (30).

Asymmetrical cleavage of the two strands

It has recently been shown (31) that the intron endonuclease I-*Ceu* I cleaves the coding strand faster than the non coding strand. Differential labeling of the two strands of synthetic substrates allowed us to test this phenomenom on the double stranded substrate with I-*Sce* III (Figure 5). We observed an effect similar to that shown for I-*Ceu* I the top (non-transcribed) strand is cleaved faster than the bottom (transcribed) strand. This observation on the wild type target sequence was extended on some mutated sequences. Four mutations, at positions −9, −2, +6, +9, that we previously found to alter the cleavage reaction, were examined. We found that they considerably slow down the cleavage of the two strands (Figure 5). In all the cases, however, the cleavage of the bottom strand remains slower than that of the top strand. This implies that the cleavage of the bottom strand is rate limiting with respect to the production of double-strand breaks.

DNA binding properties of I-*Sce* III

Very little is known about the nucleic acid binding properties of this class of proteins. However, understanding these interactions may give important clues as to their evolutionary role. Previously, we and others have failed to detect stable complexes

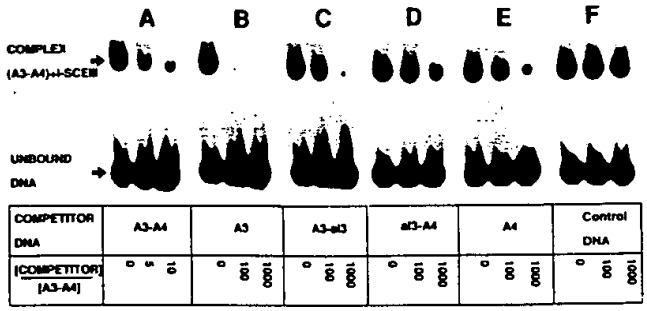


Figure 6. DNA binding activities of I-*Sce* III. Band-shift assays with purified I-*Sce* III were conducted with synthetic double-stranded DNA encompassing the A3-A4 exon fusion sequence. Competition analyses of different double stranded DNA were conducted with A) the A3-A4 sequence itself (5 and 10 fold excess), B) the upstream product 'A3' of the cleavage reaction (100 and 1000 fold excess), C) the upstream exon-intron junction (100 and 1000 fold excess), D) the downstream intron-exon junction (100 and 1000 fold excess), E) the downstream product of the cleavage reaction (100 and 1000 fold excess) and F) a completely unrelated DNA called Flag (see Materials and Methods) of similar size and composition.

between these proteins and their DNA targets. When assaying new conditions (see Materials and Methods) with this DNA endonuclease, we could observe the formation of specific DNA-protein complexes with a 20 bp long DNA target having the sequence of the A3-A4 exon junction (Figure 6). As expected for specific interactions, binding of I-*Sce* III to the 5′ labelled A3-A4 fragment could be efficiently destabilized by adding cold A3-A4 fragment. We also examined the stability of this complex in the presence of variable amounts of the products of the cleavage reaction; these two products are called A3 and A4 according to their positions relative to the exons A3 or A4. We observed that A3 efficiently competes with the full target sequence A3-A4 (Figure 6). The A4 fragment can also compete, but it is a very much weaker competitor than A3. This behaviour difference between the upstream and downstream exon sequences can also be observed, although to a lesser extent, with the upstream and downstream exon-intron junctions. To better analyze the binding properties of the cleavage products, we made two analogues of the upstream cleavage product. Both were modified in the 3′ terminal extension 5′TAAC3′ either the single stranded TAAC was deleted or it was made double strand. The original A3 DNA fragment with a single stranded protruding sequence 5′TAAC3′ is the best competitor (data not shown). If this TAAC sequence is made double-strand or if it is deleted, the corresponding A3 analogues have a lower competitor efficiency (data not shown). This suggests that the integrity of the A3 cleavage product is recognized by the enzyme. In that respect, it is important to keep in mind that all the P1-P2 DNA endonucleases characterized so far generate a 4 nucleotide 3′ extension.

DISCUSSION

In this study we have shown that the protein encoded by the third intron of the gene coding for the subunit I of cytochrome oxidase (COX I) of *Saccharomyces cerevisiae* (strain 777-3A), expressed and purified from *E.coli* is a very specific DNA endonuclease which cleaves the junction of the two flanking exons.The top strand is cleaved before the bottom strand and the protein binds differentially to upstream and downstream exonic sequences. These different points will be discussed below.

An intron-encoded DNA endonuclease I-*Sce* III

The fact that the aI3 intron-encoded protein, I-*Sce* III, has a DNA endonuclease activity, is consistent with previously known properties of the aI3 intron. We first detected, in the mitochondrial extract of a yeast strain containing the aI3 intron, a DNA endonuclease activity similar to that of the *E.coli* made protein (32) studied in this work. In particular, I-*Sce* III, like all the other known P1-P2 intron endonucleases, cleaves its target DNA and generates staggered cuts with 4-nt 3′ extensions. Moreover, the fact that these DNA endonucleases target intronless alleles of the intron-containing gene is in agreement with the recently observed intron mobility of the aI3 intron (J. Lazowska, personnal communication). The *E.coli* produced protein is 321 amino acids long, whereas the mitochondrial intronic ORF is 13 amino acids longer on the N terminal side. This intron ORF being, like most of the similar intron ORFs, in phase with the upstream exons, the mitochondrially translated protein is likely to be a chimeric exon-intron product. Whether the putative precursor is proteolysed to generate the I-*Sce* III DNA endonuclease remains to be determined. This is an interesting question in regard with the fascinating self-cleaving properties of the recently discovered homologous P1-P2 endonucleases found inserted directly in protein-coding sequences (13).

I-*Sce* III is a very specific DNA endonuclease

The second general point concerns the properties of I-*Sce* III. A common feature of the P1-P2 intron endonucleases is their specificity for long recognition sites that have an overall asymmetry. I-*Sce* III recognizes and cleaves a 20 bp long asymmetric sequence. The optimum cleavage conditions are similar to those found for the two previously purified intron endonucleases I-*Sce* I (3) and I-*Sce* II (28); in particular, there is a strict requirement for divalent cation (Mg^{2+} or Mn^{2+}). Concerning optimum pH (pH = 8) or temperature (40°C), I-*Sce* III is more similar to I-*Sce* I. The most distinctive feature of the different intron endonucleases seems to concern the degree of specifity for their target site. We found that I-*Sce* III is a very specific endonuclease. No cleavage sites were found in different genomes, such as the yeast nuclear genome, which could only be cut when a target site was artificially introduced. Also I-*Sce* III could be produced in *E.coli* without apparent effect on the cell growth. These two features are reminiscent of the properties of I-*Sce* I which is a highly specific DNA endonuclease. On the contrary we previously observed that I-*Sce* II, a related P1-P2 intron endonuclease, has a relaxed specificity. Its production was lethal to *E.coli* and, among the 18 base pairs of its target site, few bases (3 to 4) seem to be absolutely required. The mutational analysis of the I-*Sce* III target site presented in this work shows that eight mutations at different positions block cleavage completely.One may speculate on the relaxed specificity of I-*Sce* II when compared to the high specifity of I-*Sce* I (3) or I-*Sce* III (this work). In this respect it is interesting to note that, in mitochondria, the protein which corresponds to I-*Sce* II can, in different mutational contexts, acquire an RNA maturase activity (17, 33) indicating that this protein seems to be less specialized than I-*Sce* I or I-*Sce* III.

Asymmetrical cleavage and binding. Mechanistic implications

We observed, with the wild type A3-A4 target sequence, that the top strand is cleaved more rapidly than the bottom strand. A similar observation has been made with the chloroplastic I-*Ceu*

I endonuclease, suggesting that the asymmetrical cleavage is a general property of these enzymes. We examined a few mutations in the target sequence with regard to this asymmetrical cleavage of the two strands. All the mutations examined showed that the bottom strand is not cut in the absence of top strand cleavage. Taken at face value, these results suggest that cleavage of the top strand would be an essential prerequisite to cleavage of the bottom strand. This would be consistent with a two-step process for the cleavage reaction, the rate limiting step of the overall process being the cleavage of the bottom strand.

We also observed that I-*Sce* III does not bind symetrically to the upstream and downstream sequences. Systematic analyses with different synthetic DNA sequences corresponding to upstream or downstream exon sequences allowed us to show that I-*Sce* III has a greater affinity for the upstream exon cleavage product than for the downstream product. The four protruding nucleotides TAAC of the upstream product play an important role in this interaction with I-*Sce* III. It is tempting to wonder whether the asymmetrical binding of I-*Sce* III to its target might be related to the asymmetrical cleavage of the two strands. In that respect, the study of related DNA endonucleases will indicate whether there is a strict correlation between the binding properties and the fact that one strand is cleaved preferentially.

The binding of the endonuclease to the upstream and downstream exon products could, a priori, indicate three functional properties. i) A negative regulation of the cleavage activity could be carried out by the cleavage products thus limiting the action of this type of enzyme, known to be present in very low amounts in the cell. ii) The DNA endonuclease might be involved in the subsequent steps of the intron homing process by favouring the necessary heteroduplex formation. However it should be mentioned that this hypothesis does not easily fit the current view of the process. Thus, as far as T4 introns are concerned, there is good evidence that the cleavage products are degraded by exonucleases (34) before a recombinase function promotes strand invasion and D-loop formation. These last mechanistic views of the process are in agreement with the observed polarity effects also observed in the case of yeast mitochondrial intron mobility. iii) Finally, our observation that the enzyme I-*Sce* III has a higher affinity to the upstream exon might have interesting mechanistic implications. It has been reported that short sequences are conserved between the different upstream exons (18) associated, in phase, with an intron-encoded ORF. It is known that these upstream exon sequences are also involved in the formation of the P1 stem, an essential RNA element of the RNA splicing of group I intron. One could imagine that specific interaction of the protein with the upstream exon sequence could modulate either the RNA polymerase or the ribosome progression, thus controlling the production, and the folding process, of the downstream intron RNA. I-*Sce* III might have a positive role in the RNA splicing efficiency of the intron, even if it is well known that such an intron has all the properties of a self-splicing intron (14, 35). An extreme case of the development of such an ancilliary activity might be found in the well characterized RNA maturases. In that respect, knowledge of the DNA binding properties of a typical intron-encoded RNA maturase, like the bI4 RNA maturase, will be critical.

ACKNOWLEDGEMENTS

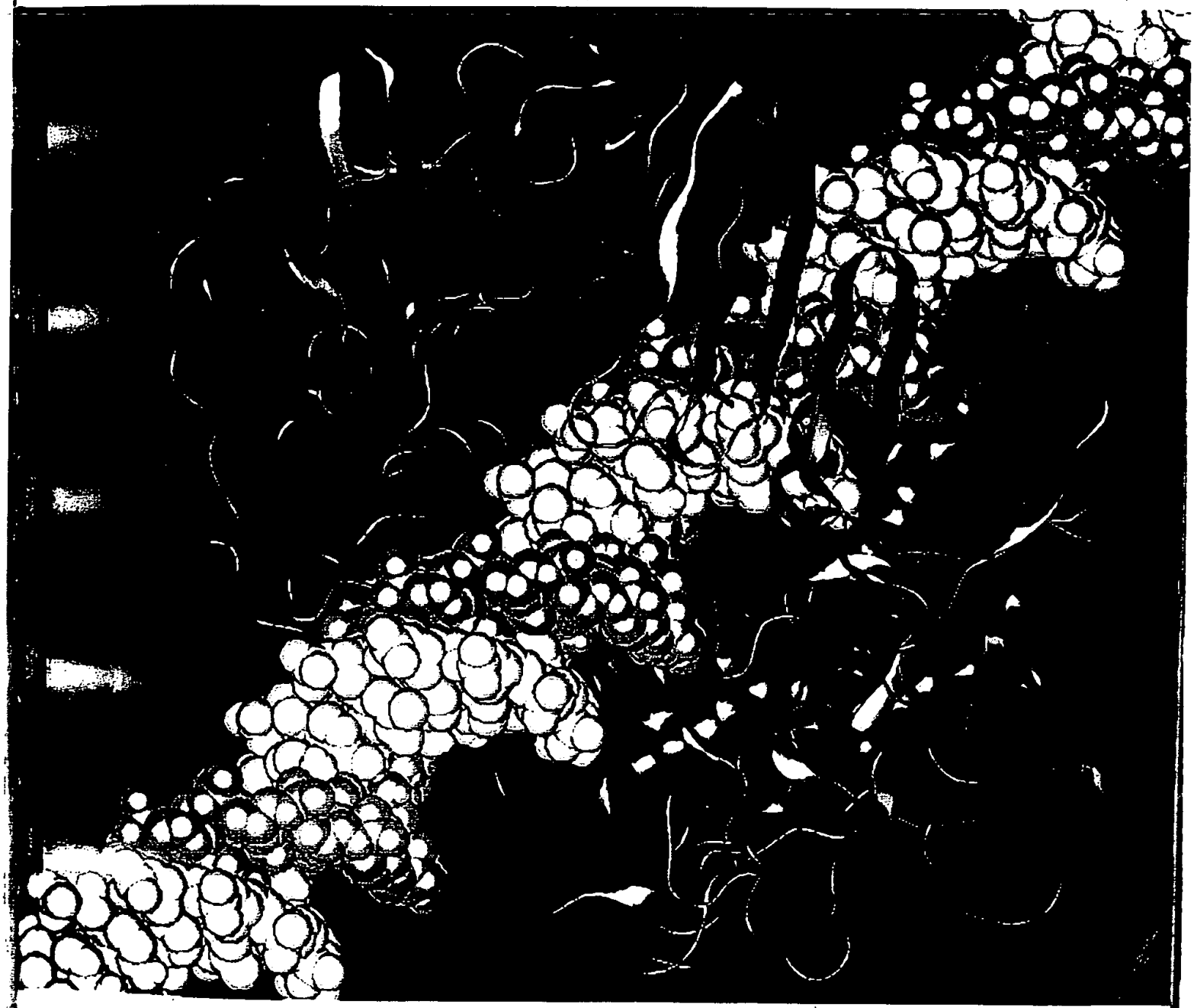
We wish to thank A.Delahodde for her wise advices and relevant suggestions all along this work and B.Sargueil for fruitful discussions. M.S. was recipient from a BDI fellowship sponsored by the CNRS and PHARMACIA.

REFERENCES

1. Lazowska, J., Jacq, C. and Slonimski, P.P. (1980) *Cell* **22**, 333–348.
2. Banroques, J., Delahodde, A. and Jacq, C. (1986) *Cell* **46**, 837–844.
3. Colleaux, L., d'Auriol, L., Betermier, M., Cottarel, G., Jacquier, A., Galibert, F. and Dujon, B. (1986) *Cell* **44**, 521–533.
4. Michel, F. and Dujon, B. (1983) *EMBO J.* **2**, 33–38.
5. Hensgens, L.A.M., Bonen, L., de Haan, M., van der Horst, G. and Grivell, L.A. (1983) *Cell* **32**, 379–389.
6. Nakagawa, K., Morishima, N. and Shibata, T. (1991) *J. Biol. Chem.* **266**, 1977–1984.
7. Nakagawa, K., Morishima, N. and Shibata, T. (1992) *EMBO J.* **11**, 2707–2715.
8. Russel, D.W., Jensen, R., Zoller, M.J., Burke, J., Errede, B., Smith, M. and Herskowitz, I. (1986) *Mol. Cell. Biol.* **6**, 4281–4294.
9. Davis, E.O., Jenner, P.J., Brooks, P.C., Colston, M.J. and Sedgwick, S.G. (1992) *Cell* **71**, 201–210.
10. Gimble, F.S. and Thorner, J. (1992) *Nature* **357**, 301–306.
11. Hirata, R. and Anraku, Y. (1992) *Biochem. Biophys. Res. Comm.* **188**, 40–47.
12. Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) *Science* **250**, 651–657.
13. Shub, D.A. and Goodrichblair, H. (1992) *Cell* **71**, 183–186.
14. Cech, T.R. (1990) *Annu. Rev. Biochem.* **59**, 543–568.
15. Lambowitz, A.M. and Perlman, P.S. (1990) *Trends Biochem. Sci.* **15**, 440–444.
16. Belfort, M. (1991) *Cell* **64**, 9–11.
17. Goguel, V., Delahodde, A. and Jacq, C. (1992) *Mol. Cell. Biol.* **12**, 696–705.
18. Dujon, B. (1989) *Gene* **82**, 91–114.
19. Bonitz, S.G., Coruzzi, G., Thalenfeld, B.E., Tzagoloff, A. and Macino, G. (1980) *J. Biol. Chem.* **255**, 11922–11926.
20. Kunkel, T.A., Roberts, J.D. and Zakovr, R.A. (1987) *Methods. Enzymol.* **154**, 367–382.
21. Oliver, S.G. et al. (1992) *Nature* **357**, 38–46.
22. Schwartz, D.C. and Cantor, C.R. (1984) *Cell* **37**, 67–75.
23. Church, G.M. and Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1996.
24. Delahodde, A., Goguel, V., Becam, A.M., Creusot, F., Perea, J., Banroques, J. and Jacq, C. (1989) *Cell* **56**, 431–441.
25. Perea, J., Desdouets, C., Schapira, M. and Jacq, C. (1993) *Nucleic Acids Res.* **21**, 358.
26. Monteilhet, C., Perrin, A., Thierry, A., Colleaux, L. and Dujon, B. (1990) *Nucleic Acids Res.* **18**, 1407–1413.
27. Wernette, C., Saldhana, R., Perlman, P. and Butow, R. (1990) *J. Biol. Chem.* **265**, 18976–18982.
28. Sargueil, B., Hatat, D., Delahodde, A. and Jacq, C. (1991) *Nucleic Acids Res.* **18**, 5659–5665.
29. Wernette, C., Saldanha, R., Smith, D., Ming, D., Perlman, P.S. and Butow, R.A. (1992) *Mol. Cell. Biol.* **12**, 716–723.
30. Thierry, A., Perrin, A., Boyer, J., Fairhead, C., Dujon, B., Frey, B. and Schmitz, G. (1991) *Nucleic Acids Res.* **19**, 189–190.
31. Marshall, P. and Lemieux, C. (1992) *Nucleic Acids Res.* **20**, 6401–6407.
32. Sargueil, B., Delahodde, A., Hatat, D., Tian, G.L., Lazowska, J. and Jacq, C. (1991) *Mol. Gen. Genet.* **225**, 340–341.
33. Dujardin, G., Labouesse, M., Netter, P. and Slonimski, P.P. (1983) *Mitochondria 1983 Nucleo-mitochondrial interactions.* 232–250.
34. Clyman, J. and Belfort, M. (1992) *Genes Dev.* **6**, 1269–1279.
35. Winter, A.J., Koerkamp, M.J.A.G. and Tabak, H.F. (1992) *Nucleic Acids Res.* **20**, 3897–3904.

VOLUME 21 NUMBER 16 AUGUST 11, 1993

Nucleic Acids Research



IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

ISSN 0305 1048 Coden NARHAD